# Supplementary Information

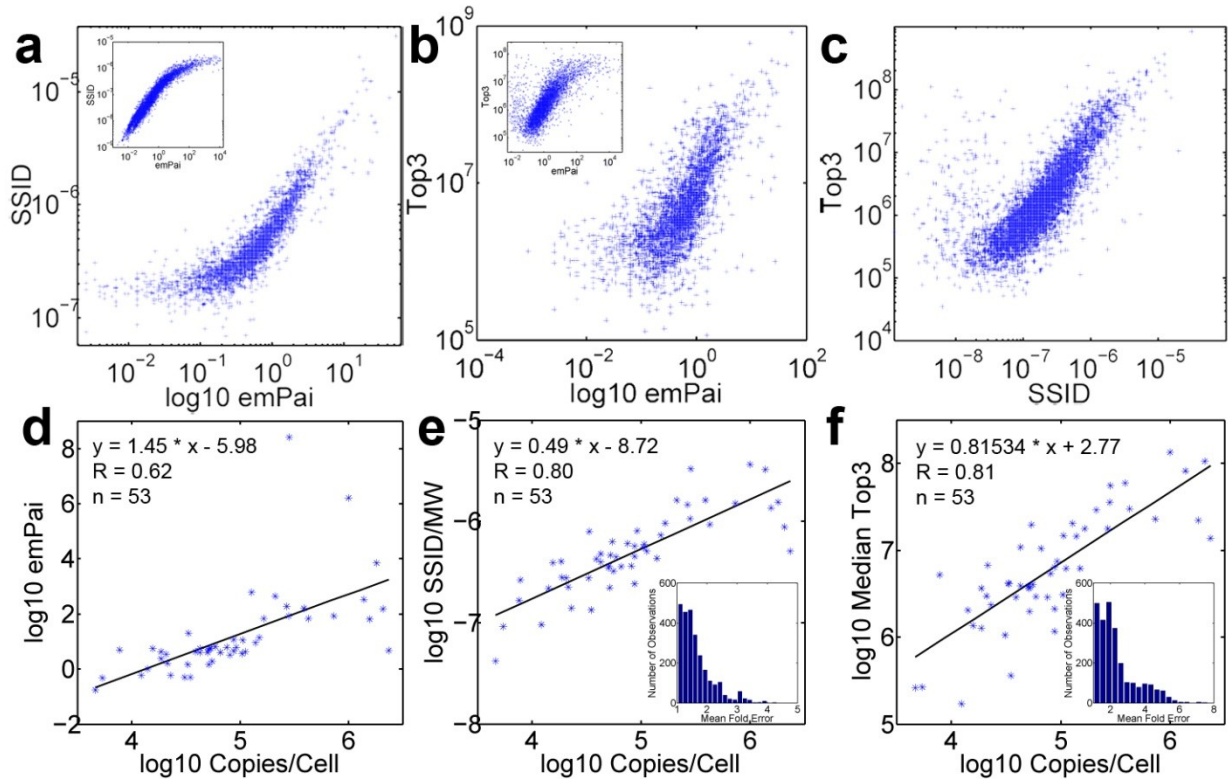


**Figure S1: Comparison different protein abundance indices and calibration to protein copies per cell.** **(a)** The log10 of emPai is plotted against SSID index (log-scale) shows a correlation over three orders of magnitude but not below emPai values of 1e-1. Inset: emPai against SSID (log-scale). **(b)** Log-scale plot of log10 emPai against Top3. Inset: emPai against Top3 (log-scale). **(c)** The log-scale plot of the SSID against the Top3 indices show a correlation over more than three orders of magnitude. **(d)** Double-logarythmic plot of copy numbers per cell of a selected subset of proteins as determined using AQUA peptides against the emPai. **(e)** The same for the SSID index. Inset: Error model as determined by bootstrapping analysis (Number of iterations = 2500, sample size = 3, Mean Fold Error ~ 1.7). **(f)** The same for the Top3 index. Inset: as in (e), Mean Fold Error ~ 2.4.

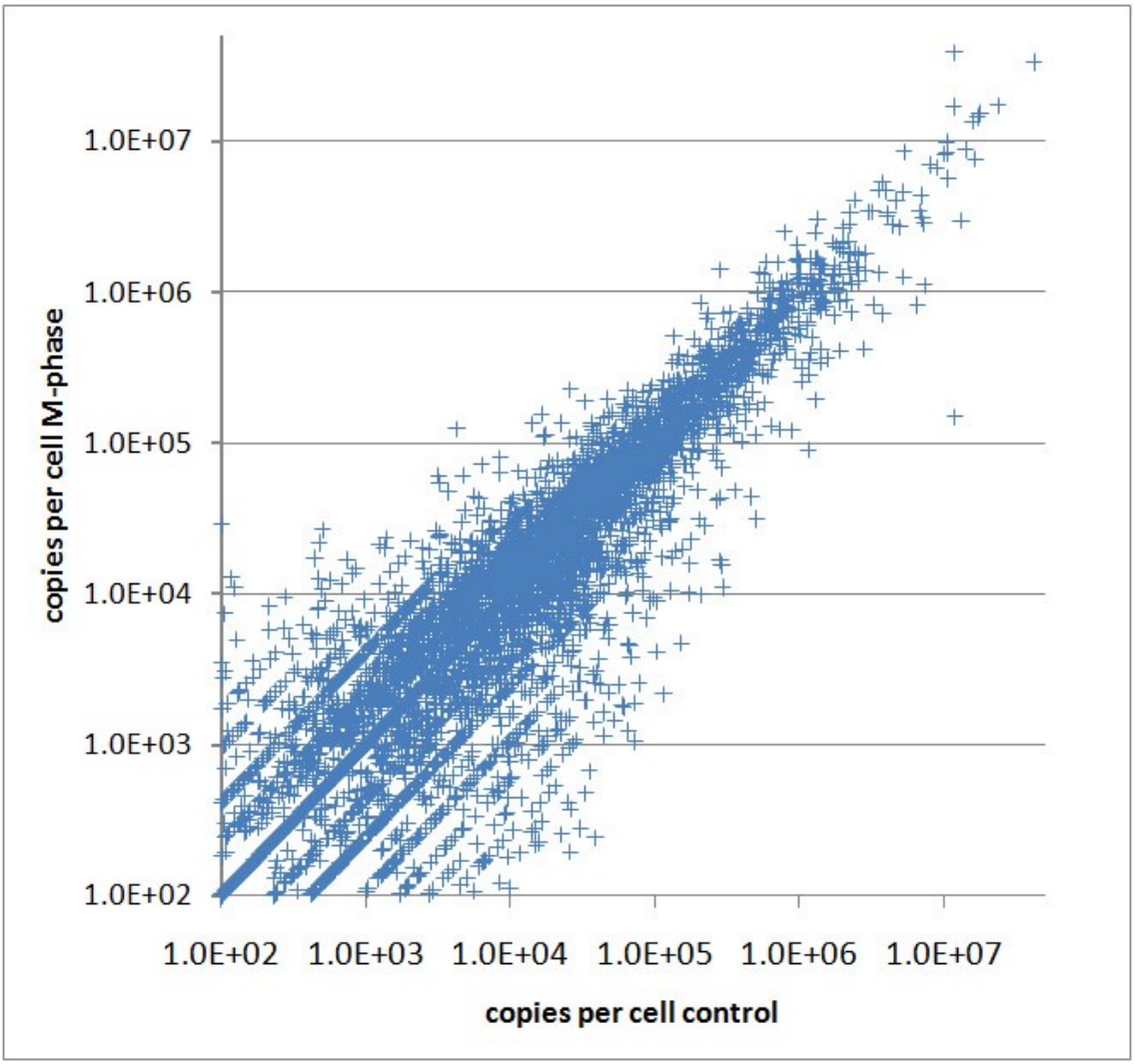


**Figure S2. Logarithmic scale plot of the protein copies per cell measured in non-synchronized and nocodazole-treated cells.** The displayed primary data have not been masked from 5e2 to 2e7 copies due to precision confidence as done for Table S1.

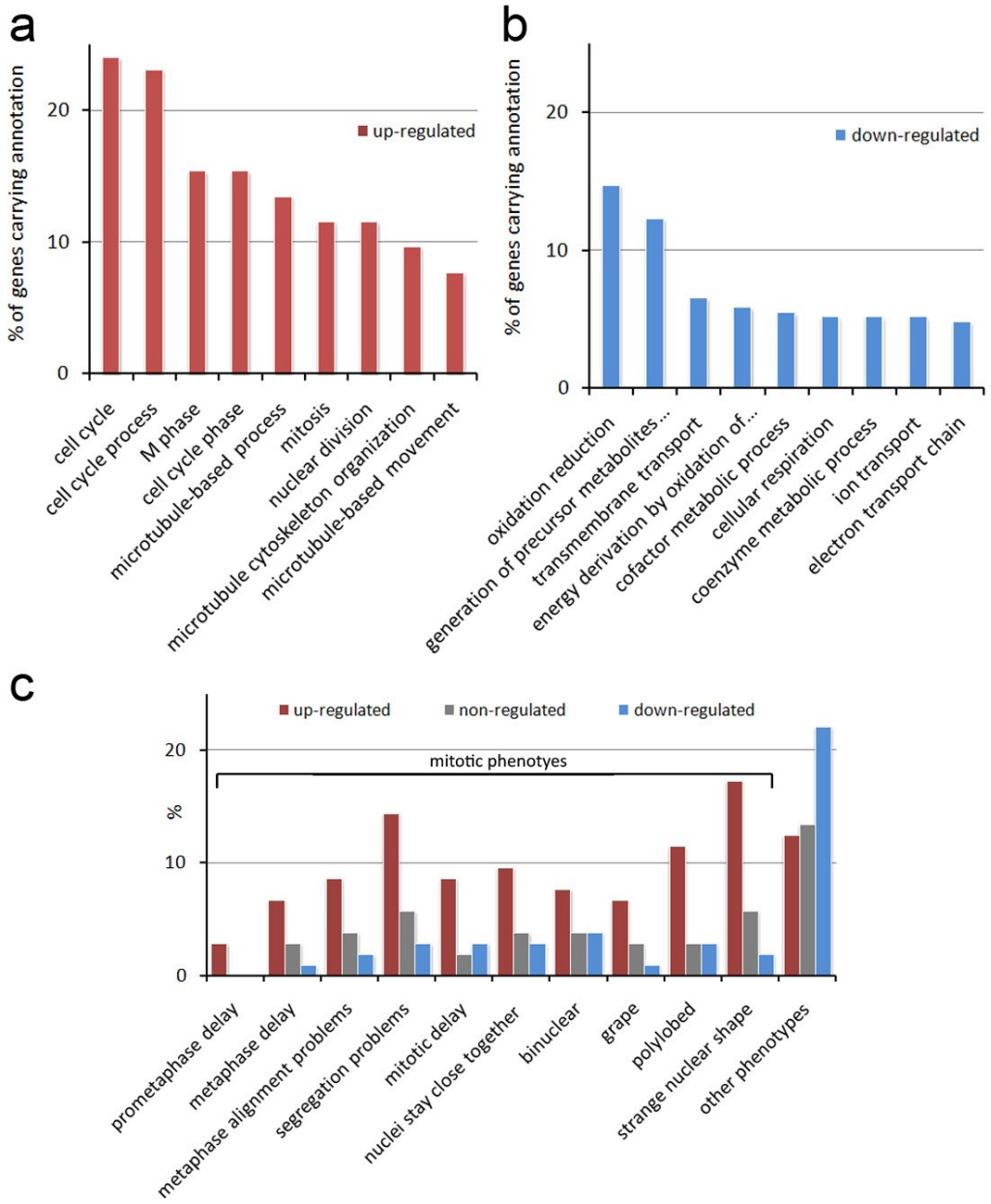


**Figure S3. Functional annotations of proteins showing copy number variations during M-phase.** Proteins found significantly up-regulated (a) and down-regulated (b) during Mitosis as compared to all detected proteins are shown. The most enriched functional categories among up-regulated proteins are mitotic processes, while among down-regulated proteins metabolic processes are enriched (Benjamini-corrected *p* value of <1e-1 as confidence threshold for enrichment of GO-categories). (c) Proteins that were found up-regulated during mitosis (red) are enriched for mitotic phenotypes after gene knock down as compared to non- and down-regulated proteins (grey, blue; each gene can have multiple phenotypes). The three protein groups were compared to mitotic phenotypes discovered during a recent genome-wide gene silencing study (Neumann et al, 2010). The column to the right shows the sum of all phenotypes categorized as unrelated to mitosis (including cell death and migration, which are not displayed individually).

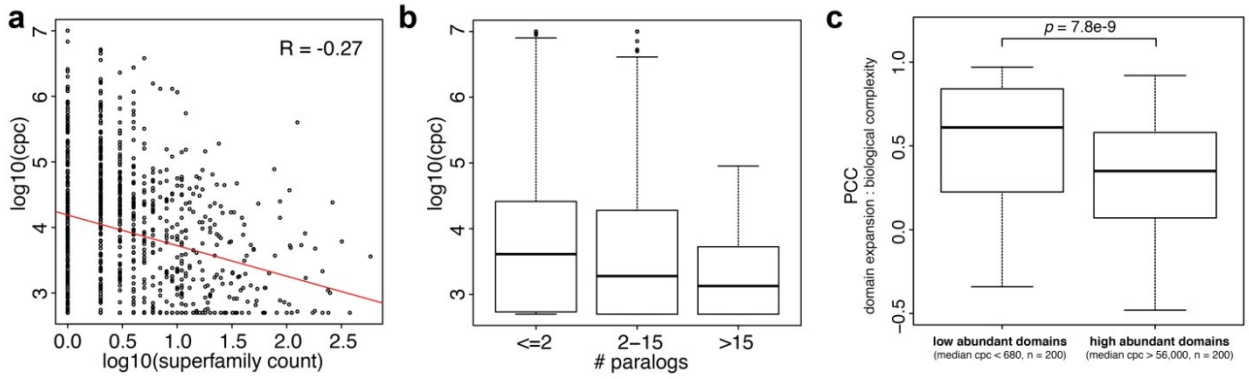


**Figure S4. Protein abundance affects domain expansion and gene duplication during evolution.** The frequency of superfamilies in the human genome (a) and the number of paralogs per protein (b) have a negative correlation with protein abundance. (c) The expansion of domains expressed at low copy number has a stronger correlation with increase in organism complexity than the one of abundant domains. The correlation coefficients between superfamily expansion and increase in biological complexity were obtained from (Vogel & Chothia, 2006). The correlation coefficients for the 200 most abundant (cpc > 56,000) and 200 least abundant (cpc < 680) superfamilies were compared (Table S5). *p* value was calculated using a one-sided Wilcoxon rank sum test. PCC: Pearson Correlation Coefficient; cpc: copies per cell.

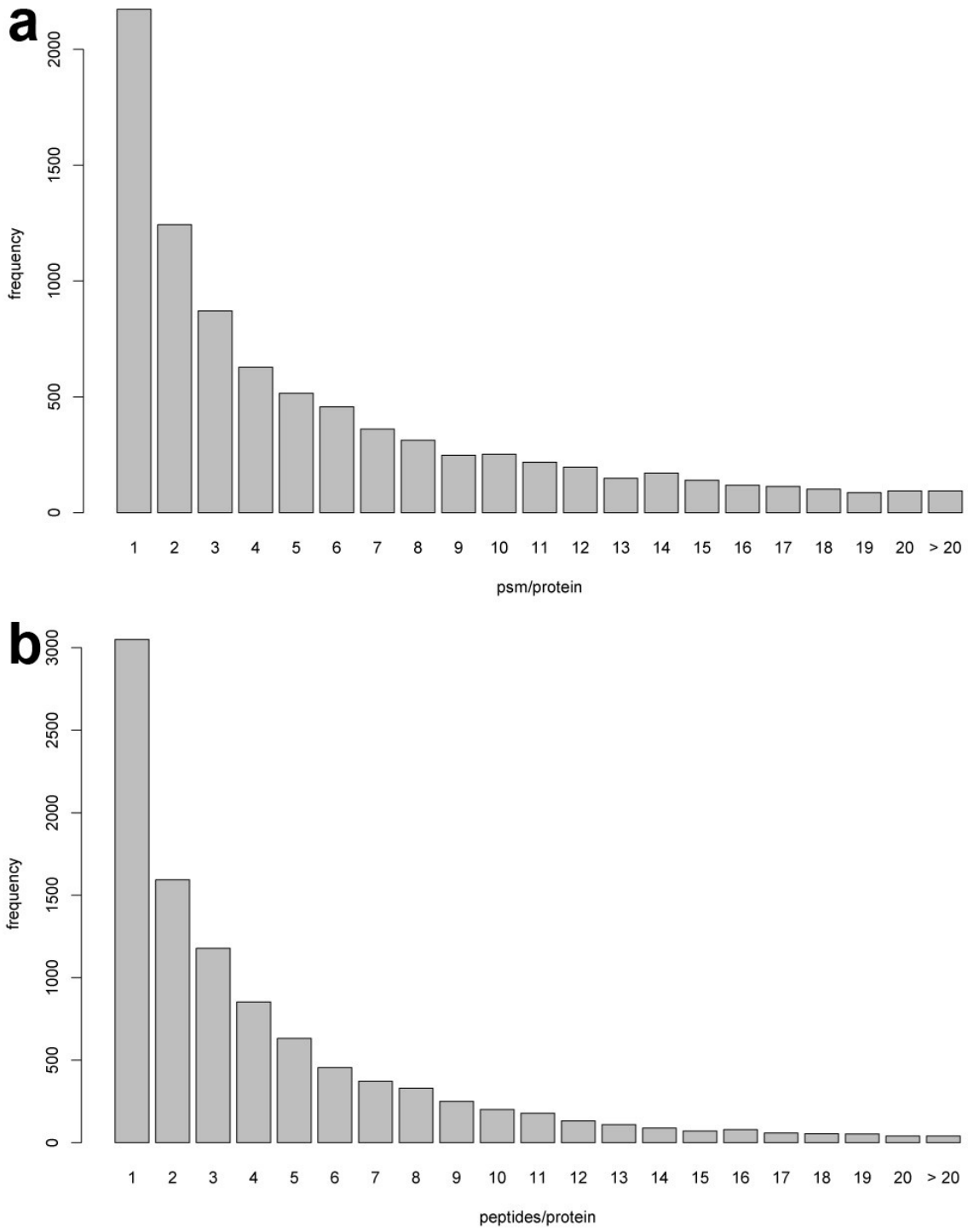


**Figure S5. Number of peptide spectra matches (a) and peptides (b) per protein.** The protein false positive discovery rate was 1.2% and was calculated using the Mayu software. The respective set of PSM achieving this protein FDR has a PSM FDR of 0.14 %.

**Supplementary Methods**

**Cell culture and synchronization.** U2OS cells were obtained from the American type culture collection (order number HTB-96), grown in DMEM medium supplemented with 10% Bovine Serum, harvested by trypsinization, washed twice in Phosphate Buffered Saline (PBS) and counted using a Neubauer chamber (n = 10, mean count = 41, std = 4.9). 1e7 cells were resuspended in 0.5 ml PBS and lysed by adding 0.5 ml of 10mM Tris pH 7.5, 10M urea, 0.1% Rapigest immediately before sonication for 5 min. Lysates were cleared by centrifugation at 13.000 rpm for 10 min in a table-top Eppendorf centrifuge. Synchronization was carried out by adding nocodazole to a final concentration of 330 nM for 18 hours and harvesting through 'shake-off'.

**Mass Spectrometry.** Proteins were reduced with 10 mM TCEP for 20 min at 37°C and alkylated with 10 mM iodoacetamide for 20 min in the dark at room temperature before diluting the sample with 100 mM ammonium bicarbonate to a final urea concentration below 1.5 M. Proteins were digested by incubation with trypsin (1/100, w/w) for at least 6 hours at 37°C. AQUA peptides were spiked into the sample at this stage, if applicable. The peptides were cleaned up by C18 reversed-phase spin columns according to the manufacturer's instructions (Harvard Apparatus). For Off-gel electrophoresis, the dried down peptides were re-solubilized to a final concentration of 1 mg/ml in off-gel electrophoresis buffer containing 6.25% glycerol and 1.25% IPG buffer (GE Healthcare). The peptides were separated on pH 3-10 IPG strips (GE Healthcare) with a 3100 OFFGEL fractionator (Agilent) according to the manufacturer's instructions using a protocol of 1 hour rehydration at maximum 500 V, 50 µA and 200 mW followed by the separation at maximum 8000V, 100 µA and 300 mW until 50 kVh were reached. After iso-electric focusing all 22 fractions were concentrated and cleaned up by C18 reversed-phase spin columns according to the manufacturer's instructions (Harvard Apparatus). After an initial LC-MS analysis, neighboring low complexity fractions were pooled to a final of 16 fractions. The setup of the µRPLC-MS system was as described previously (Schmidt et al, 2008) with some modifications. The hybrid LTQ-FT-ICR mass spectrometer was interfaced to a nanoelectrospray ion source (both ThermoScientific, Bremen, Germany) coupled online to a Tempo 1D-plus nanoLC (Applied Biosystems/MDS Sciex, Foster City, CA). 1 µg of peptides were separated on a RP-LC column (75 µm x 15 cm) packed in-house with C18 resin (Magic C18 AQ 3 µm; Michrom BioResources, Auburn, CA, USA) using a linear gradient from 96 % solvent A (98 % water, 2 % acetonitrile, 0.15 % formic acid) and 4% solvent B (98 % acetonitrile, 2 % water, 0.15 % formic acid) to 30 % solvent B over 120 minutes at a flow rate of 0.3 µl/min. Each survey scan acquired in the ICR-cell at 100,000 FWHM was followed by MS/MS scans of the three most intense precursor ions in the linear ion trap with enabled dynamic exclusion for 60 seconds. Charge state screening was employed to select for ions with at least two charges and rejecting ions with undetermined charge state. The normalized collision energy was set to 32% and one microscan was acquired for each spectrum. Samples comparing different cell cycle states were analyzed using an Easy-nLC / Orbitrap-Velos (both ThermoScientific, Bremen, Germany) LC-MS system with the following modified parameters; Peptides were separated using a linear gradient from 92 % solvent A (98 % water, 2 % acetonitrile, 0.15 % formic acid) and 8% solvent B (98 % acetonitrile, 2 % water, 0.15 % formic acid) to 40 % solvent B over 120 minutes. Each survey scan acquired in the Orbitrap at 60,000 FWHM was followed by MS/MS scans of the 20 most intense precursor ions in the linear ion trap.

Depending on the sample complexity, each fraction was analyzed 3-4 times in shotgun and 2-5 times in directed (inclusion list) mode. Directed LC-MS measurements of features and reference peptides were performed according to (Schmidt et al, 2008) using a rolling inclusion list if the number of masses exceeded 500. Inclusion lists were generated as follows: Data from the 74 MS runs carried out in shot gun mode were subjected to SuperHirn analysis to obtain peak intensities of all features from extracted ion chromatograms (XICs) for each OGE fraction. The entity of all feature intensities together with their elution time, charge state and matching MS2 spectra is thereby stored in a so-called MasterMap (Mueller et al, 2007). In order to systematically target all the identified XICs that did not match to the

already acquired MS2 spectra we designed software termed 'inclusion list generator'. This software parses the relevant MasterMaps, identifies all non-sequenced XICs ions in a OGE fraction and distributes them into scheduled inclusion list with respect to the scanning speed of the mass spectrometer. Thereby, large feature lists were automatically split into smaller lists covering a certain mass range and charge to enable more specific directed mass spectrometry analysis (Scherl et al, 2008). Inclusion list generator suggested another 53 scheduled inclusion list runs. These 53 runs were carried out as described above but in 'inclusion list'-mode (Schmidt et al, 2009) employing gas phase and charge state fractionation when necessary.

**MS/MS data processing.** After converting the acquired raw files to the centroid mzXML format (readW), MS/MS spectra were searched using the SEQUEST algorithm (Yates et al, 1995) against a decoy database (consisting of forward and reverse protein sequences) of all non-redundant human Swiss Prot and Trembl entries V57.0 (proteome mapping experiment) or of all human Swiss Prot entries V 57.12 (quantification experiments), respectively, and known contaminants such as porcine trypsin, human keratins and high abundant bovine serum proteins. The search criteria were set as follows: full tryptic specificity was required (cleavage after lysine or arginine residues, unless followed by proline); 2 missed cleavages were allowed; carbamidomethylation (C) was set as fixed modification; oxidation (M), 13C6-15N2 (K) and 13C6-15N4 (R) were applied as variable modifications; mass tolerance of 15 ppm (precursor) and 0.8 Da (fragments). The database search results were further processed using the PeptideProphet (Keller et al, 2002) and ProteinProphet (Keller et al, 2005) program.

**Data deposition.** The three MS raw data sets corresponding to the proteome mapping and both quantification experiments (of synchronized and non-synchronized cells) were deposited at Proteome Commons (https://proteomecommons.org) with the following hash codes:

3Wj0424JA2DCVkBnfqm45v+UfMZOHgf3p2PTwUe83RwjqQvtr4mQnloYvUSrMHCYBz+krDIXmz50spF2TN

YGw3/8jZIAAAAAAAAB9w==

x8hmYUs40bOspaY+EMuyUtDkyiw+xgyjSynVK/ggQXhl+bbDV5QbiAMakzsSKonz/XszxEEUThtmn6cIS/STS1
Y0n2QAAAAAAAAB3g==

Gm5TsXK3crQV70MqiIIH+/uaKyioNCFWi+Ri7fpLq+W1ga5OQA0dTe2u0LMvN+ty7uuRsA1o3WTWb79Bc/
XqYK7v9D0AAAAAAAACBA==

**FDR Estimation.** Peptide-spectrum match (PSM) false discovery rates have been estimated by means of the target-decoy strategy (Elias & Gygi, 2007). Protein false discovery rates have been estimated by the generalized target-decoy strategy (Reiter et al, 2009). Briefly, protein false discovery rates are estimated from the number of decoy and target protein discoveries by means of a hypergeometric model. Naive protein false discovery rates have been estimated by taking the ratio of decoy and target protein discoveries as proposed earlier (Nesvizhskii et al, 2007). The resulting distribution of peptides and PSMs per protein is shown in Figure S5. The naive estimate computes protein FDR directly from the number of decoy protein identifications, thereby neglecting the situation of true protein identifications being both supported by true and false peptide spectrum matches. Mayu corrects for this phenomenon by appropriately modeling its statistical implications on the FDR estimate. This correction is particularly important in deep proteome sequencing projects, where this type of chimeric protein identifications occur frequently.

**Proteome Coverage Prediction.** Proteome coverage prediction has been performed as described in (Claassen et al, 2011). Briefly, we assume a generalized hierarchical Pitman-Yor process prior for the ensemble of peptide distributions arising and sampled in the course of the many different LC-MS/MS experiments carried out to generate the "u2os dataset". We compute an empirical Bayes parameters estimate for the process priors and simulate further experiments to estimate the expected number of protein discoveries.

**Generation and of global MasterMaps.** To calculate the Top3 index from the entity of all XICs in all measured OGE fractions and to relatively quantify cells from the control condition versus the synchronized condition, global MasterMaps were calculated as follows: XICs from all MS runs were calculated using Progenesis (nonlinear dynamics). All XICs were parsed into Matlab (The MathWorks), all measured features accounting for the same peptide were summed up over all OGE fractions and charges states using Matlab. This union took into account the extracted ion currents (XICs) observed for each peptide species at all possible charge states in all OGE fractions but kept sequence variants of these peptides, e.g. peptides carrying post-translational modifications, as separate species.

**Absolute Abundance Estimation.** AQUA peptides were grouped into three abundance classes based on spectral counts obtained for the corresponding endogenous peptides and spiked into human peptide mixtures directly after digestion (see above), at a final concentration of 0.5, 5 or 50 pmol/μL, respectively. Heavy and light ratios between spiked AQUA and endogenous peptides were calculated using XPRESS as implemented into the trans-proteomic pipeline (Keller et al, 2005). After excluding low confidence assignments of which heavy and light pairs were observed independently less than three times or that had MS2 spectra only for either the light or heavy species, ratios for 71 peptides corresponding to 53 proteins were obtained.

We used this quantitative data to calculated protein abundance indices for the identified U2OS proteins by three different methods: the exponentially modified protein abundance index (emPAI) was calculated as previously described (Ishihama et al, 2005) and based on the peptide statistics calculated by PeptideProphet (Keller et al, 2002). To the share of spectrum identification index (SSID) the 'percent share of spectrum ID' output of ProteinProphet (Keller et al, 2005) was normalized by dividing through the proteins molecular weight. The median of the extracted precursor ions (XICs) of the three best flying peptides per protein (Top3) index (Silva et al, 2006) was calculated as described previously (Malmstrom et al, 2009) but based on XICs calculated using the Progenesis software (nonlinear dynamics). Of these, only the latter index requires the global MasterMap as input. All three indices yielded similar absolute abundance values on an arbitrary scale over several orders of magnitude (Figure S1 a-c). The Top3 and SSID indices correlated relatively well over approximately 4 orders of magnitude of protein concentrations (Fig S1b, c). To calibrate the arbitrary protein abundance indices to absolute protein copy numbers per cell we determined the concentration of the subset of endogenous light peptides corresponding to the spiked-in heavy labeled reference peptides (Han et al, 2001). We confidently detected and quantified 70 reference and endogenous peptides pairs corresponding to 53 proteins in a concentration range from 4.5e3 to 2.5e6 copy numbers per cell. We next validated the precision of the abundance indices by bootstrapping analysis against the absolutely quantified protein set as described earlier (Malmstrom et al, 2009). The correlation between the emPai index and protein copy numbers was only moderate in our data, while the SSID and Top3 indices showed acceptable correlations ($R > 0.8$). The mean fold error of these two indices was about 1.7 and 2.4, respectively (Figure 2e, f insets) (See supplement for detail). The Top3 index performed slightly worse than SSID in the mammalian cells, in contrast to the bacterial system (Malmstrom et al, 2009) where it outperformed the SSID. This might be explained with the additional sample fractionation steps required for the analysis of the more complex human samples which complicated accurate feature extraction. Another reason for the slightly

lower performance of the Top3 index could be the smaller number of detectable proteotypic peptides per protein in mammals.

The precision (mean fold error) of absolute protein abundance indices was estimated using bootstrapping analysis as described before (Malmstrom et al, 2009). In brief, protein abundances absolutely measured using AQUA (n = 53) were correlated with the relevant protein abundance index on log-log-scale. A random subset of proteins (n = 3) was removed from the main set and the error for the subset calculated after rebuilding the linear model with the remaining proteins. The procedure was repeated 2500 to calculate a robust mean fold error for each abundance index. The linear model was afterwards used to calculate the absolute abundance of all proteins covered by the the SSID abundance index (emPAI and Top3 displayed a weaker performance or were covering fewer proteins).

To estimate the average number of nuclear pore complexes per cell we performed immunofluorescence labeling, high resolution confocal microscopy and computational image analysis. In brief, U2OS cells were pre-permeabilized with 0.1% Triton X-100 for 3 min, fixed with 2% paraformaldehyde in PBS for 15 min and permeabilized with 0.1% Triton X-100 for an additional 10 min. NPCs were then stained with mab414 (Covance) and secondary goat anti-mouse IgG coupled to Alexa Fluor 488 (Invitrogen). DNA was visualized by Hoechst 33342 (Sigma). Confocal z-stacks (xyz pixel size 44/44/380 nm) through the entire nucleus of 46 randomly chosen cells were acquired using a Zeiss LSM 710 confocal microscope with Plan-Apochromat 63x/1.4 Oil DIC objective. In each cell the NPC density was quantified on an area representing approximately 1/3 of the lower nuclear membrane in the Alexa 488 channel with an in-house developed intensity peak identification macro in Image J. The total number of pores per cell was extrapolated by multiplying the density with the surface of the nucleus obtained after segmentation and 3D reconstruction of the Hoechst 33342 channel using Imaris x64 6.4.0 (Bitplane). To estimate the number of NPCs from MS data, copy numbers per cell of all detected nucleoporins were divided by their copy number per NPC. For simplicity, 16 copies per NPC were assumed for all components which are within the tolerable error of the MS method. Transmembrane domain containing proteins were omitted since they were generally underrepresented in the proteome (Table S4).

Synchronized cells were analyzed as described above and the absolute abundance scale of proteins during mitosis was established using the same model based on the mitotic SSID index. By comparing the SSID indices from the nocodazole-arrested and non-synchronized cell data sets we quantified more than 6,100 proteins across both cellular states (Table S2). The detection of protein abundance changes from the SSID index is expected to be less accurate because it is not based on multiple peptide measurements but individual signals. To test weather proteins regulated during mitosis are associated with known mitotic phenotypes as identified by RNAi screening (Figure S3) we carried out relative, label-free quantification as follows: A second global MasterMap fro al samples of synchronized cells was generated and peptide ratios between the two global MasterMaps were calculated. Afterwards protein ratios were calculated if at least two proteotypic peptides per protein were observed in both data sets. As additional quality criteria, peptides were not considered if they did not elute within a retention time window of 10 min in both data sets and if their ratio was not within a 1-times standard deviation window of the average ratio of all peptides of the same protein. We confidently quantified more than 4,000 proteins across both states and detected 104 proteins that were significantly up- and 293 proteins that were significantly down-regulated during M-phase (Table S2). Proteins were considered to be regulated if their ratio was outside of 3-sigma significance threshold, accounting for being more than 2.5-fold regulated.

**Functional analysis.** The functional annotation of the U2OS proteome was performed using a custom designed GO slim annotation. GO slims are reduced versions of the GO ontologies that contain only a subset of terms in order to provide a high level view of the ontology content of a dataset (Rhee et al, 2008). Thus, a set of broad GO biological process terms (Table S6) were selected using QuickGO (annotation revision version 5.1331) (Binns et al, 2009) and used to annotate the entries included in the

reference database used for MS/MS spectra search. 14,887 of the 20,329 (73%) gene models present in the reference database were annotated with at least a biological process term. The GO slim annotation was used to calculate the fraction of annotations for the different biological processes at the genome level (relatively to the total number of annotations derived from the gene models in the reference database), qualitative proteome level (relatively to the total number of annotations derived from the detected proteins) and quantitative proteome level (relatively to the sum of protein copies for all the annotations). The average number of annotations was 1.7 per gene model and 1.9 per detected protein. We also investigated the abundance distribution of selected protein classes (Figure 2b,c). The protein classes were defined using a combination of GO terms (Table S7) on the basis of their belonging to protein complexes (ribosome and proteasome), specific molecular functions (transmembrane receptors, transcription factors, kinases and phosphatases) or involvement in defined biological processes (protein folding, glycolysis and cell adhesion).

The under-/over-representation of GO cellular component terms in the detected proteome (Table S4) was analyzed with the Cytoscape plugin BiNGO v2.44 (Maere et al, 2005; Shannon et al, 2003), using the whole human annotation as reference set, by applying a hypergeometric test with Benjamini-Hochberg false discovery rate correction (Benjamini & Hochberg, 1995). The enrichment of PANTHER terms (Thomas et al, 2003) and KEGG pathways in different abundance classes (Table S3) was analyzed with DAVID 6.7 (Huang da et al, 2009) using the quantified proteome as a background list (n=7309). For the enrichment of GO biological process terms in proteins regulated in synchronized cells, the set of proteins quantified across the two cell states (n=4,036 for MasterMaps and n=6,164 for SSID) (Table S2) was used as background list. The enrichments were statistically analyzed using a modified Fisher's exact test (EASE score) (Huang da et al, 2009) with Benjamini-Hochberg false discovery rate correction. *p* value cut-offs were set at 0.05 unless otherwise stated.

The comparative analysis of quantitative proteomes across species (Figure 3) was performed using published dataset for *L.interrogans* (Malmstrom et al, 2009), *S.cerevisiae* (Ghaemmaghami et al, 2003) and mouse (Schwanhausser et al, 2011), and the dataset presented in this study for human. All the protein identifiers were mapped to UniProt/SwissProt accession numbers. After conversion, the quantitative dataset covered 49.6% (1,816/3,654), 57% (3,815/6,696), 23% (4,962/21,873) and 35% (7,309/20,599) of the known protein-coding genes for *L.interrogans, S.cerevisiae*, mouse and human, respectively. The number of known protein-coding genes were retrieved from Ensembl (Flicek et al, 2011) for *S.cerevisiae* (assembly EF 3, Feb 2011), mouse (assembly NCBIM37, Apr 2007) and human (assembly GRCh37.p3, Feb 2009), and from the same source as decribed in (Malmstrom et al, 2009) for *L.interrogans*. GO annotations were downloaded from QuickGO (annotation revision version 5.1331) using the slim terms listed in Table S6. Not annotated proteins were 42%, 11%, 14% and 18%, while the average number of annotations per quantified protein were 1.3, 1.7, 1.9 and 1.9 for *L.interrogans, S.cerevisiae*, mouse and human, respectively.

**Protein evolution.** Domain distribution across quantified proteins was investigated using the Superfamily v1.75 annotation (*Uniprot* 2011_07 assignment) (Gough et al, 2001). In total, 76% (5,589/7,309) of the quantified proteins were assigned to at least one superfamily. The average number of annotation per protein was 2. The number of paralogs for human genes was downloaded from Ensembl (assembly GRCh37.p3, Feb 2009).

**Table S6**

| GO term | Term description |
|---|---|
| GO:0006139 | nucleobase, nucleoside, nucleotide and nucleic acid metabolic process |
| GO:0006810 | transport |
| GO:0019538 | protein metabolic process |
| GO:0016310 | phosphorylation |
| GO:0006350 | transcription |
| GO:0023052 | signaling |
| GO:0006412 | translation |
| GO:0005975 | carbohydrate metabolic process |
| GO:0006260 | DNA replication |
| GO:0006629 | lipid metabolic process |
| GO:0032502 | developmental process |
| GO:0007154 | cell communication |
| GO:0007155 | cell adhesion |
| GO:0007010 | cytoskeleton organization |

**Table S7**

| Functional group | GO term | Term description |
|---|---|---|
| Ribosome | GO:0022626 | cytosolic ribosome |
| Proteasome | GO:0000502 | proteasome complex |
| Protein folding | GO:0006458 | 'de novo' protein folding |
| Glycolysis | GO:0006096 | glycolysis |
| Transmembrane receptors | GO:0004888 | transmembrane receptor activity |
| Transcription factors | GO:0001071 | nucleic acid binding transcription factor activity |
| Kinases | GO:0016301 | kinase activity |
| Phosphatases | GO:0016791 | phosphatase activity |
| Cell adhesion | GO:0007155 | cell adhesion |

**Supplementary References**

Benjamini Y, Hochberg Y (1995) Controlling the False Discovery Rate - a Practical and Powerful Approach to Multiple Testing. *J Roy Stat Soc B Met* **57:** 289-300

Binns D, Dimmer E, Huntley R, Barrell D, O'Donovan C, Apweiler R (2009) QuickGO: a web-based tool for Gene Ontology searching. *Bioinformatics* **25:** 3045-3046

Claassen M, Aebersold R, Buhmann JM (2011) Proteome coverage prediction for integrated proteomics datasets. *J Comput Biol* **18:** 283-293

Elias JE, Gygi SP (2007) Target-decoy search strategy for increased confidence in large-scale protein identifications by mass spectrometry. *Nat Methods* **4:** 207-214

Flicek P, Amode MR, Barrell D, Beal K, Brent S, Chen Y, Clapham P, Coates G, Fairley S, Fitzgerald S, Gordon L, Hendrix M, Hourlier T, Johnson N, Kahari A, Keefe D, Keenan S, Kinsella R, Kokocinski F, Kulesha E et al (2011) Ensembl 2011. *Nucleic Acids Res* **39:** D800-806

Ghaemmaghami S, Huh WK, Bower K, Howson RW, Belle A, Dephoure N, O'Shea EK, Weissman JS (2003) Global analysis of protein expression in yeast. *Nature* **425:** 737-741

Gough J, Karplus K, Hughey R, Chothia C (2001) Assignment of homology to genome sequences using a library of hidden Markov models that represent all proteins of known structure. *J Mol Biol* **313:** 903-919

Han DK, Eng J, Zhou H, Aebersold R (2001) Quantitative profiling of differentiation-induced microsomal proteins using isotope-coded affinity tags and mass spectrometry. *Nat Biotechnol* **19:** 946-951

Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4:** 44-57

Ishihama Y, Oda Y, Tabata T, Sato T, Nagasu T, Rappsilber J, Mann M (2005) Exponentially modified protein abundance index (emPAI) for estimation of absolute protein amount in proteomics by the number of sequenced peptides per protein. *Mol Cell Proteomics* **4:** 1265-1272

Keller A, Eng J, Zhang N, Li XJ, Aebersold R (2005) A uniform proteomics MS/MS analysis platform utilizing open XML file formats. *Mol Syst Biol* **1:** 2005 0017

Keller A, Nesvizhskii AI, Kolker E, Aebersold R (2002) Empirical statistical model to estimate the accuracy of peptide identifications made by MS/MS and database search. *Anal Chem* **74:** 5383-5392

Maere S, Heymans K, Kuiper M (2005) BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. *Bioinformatics* **21:** 3448-3449

Malmstrom J, Beck M, Schmidt A, Lange V, Deutsch EW, Aebersold R (2009) Proteome-wide cellular protein concentrations of the human pathogen Leptospira interrogans. *Nature* **460:** 762-U112

Mueller LN, Rinner O, Schmidt A, Letarte S, Bodenmiller B, Brusniak MY, Vitek O, Aebersold R, Muller M (2007) SuperHirn - a novel tool for high resolution LC-MS-based peptide/protein profiling. *Proteomics* **7:** 3470-3480

Nesvizhskii AI, Vitek O, Aebersold R (2007) Analysis and validation of proteomic data generated by tandem mass spectrometry. *Nat Methods* **4:** 787-797

Neumann B, Walter T, Heriche JK, Bulkescher J, Erfle H, Conrad C, Rogers P, Poser I, Held M, Liebel U, Cetin C, Sieckmann F, Pau G, Kabbe R, Wunsche A, Satagopam V, Schmitz MH, Chapuis C, Gerlich DW, Schneider R et al (2010) Phenotypic profiling of the human genome by time-lapse microscopy reveals cell division genes. *Nature* **464:** 721-727

Reiter L, Claassen M, Schrimpf SP, Jovanovic M, Schmidt A, Buhmann JM, Hengartner MO, Aebersold R (2009) Protein identification false discovery rates for very large proteomics data sets generated by tandem mass spectrometry. *Mol Cell Proteomics* **8:** 2405-2417

Rhee SY, Wood V, Dolinski K, Draghici S (2008) Use and misuse of the gene ontology annotations. *Nat Rev Genet* **9:** 509-515

Scherl A, Shaffer SA, Taylor GK, Kulasekara HD, Miller SI, Goodlett DR (2008) Genome-specific gas-phase fractionation strategy for improved shotgun proteomic profiling of proteotypic peptides. *Anal Chem* **80:** 1182-1191

Schmidt A, Claassen M, Aebersold R (2009) Directed mass spectrometry: towards hypothesis-driven proteomics. *Curr Opin Chem Biol* **13:** 510-517

Schmidt A, Gehlenborg N, Bodenmiller B, Mueller LN, Campbell D, Mueller M, Aebersold R, Domon B (2008) An integrated, directed mass spectrometric approach for in-depth characterization of complex peptide mixtures. *Mol Cell Proteomics* **7:** 2138-2150

Schwanhausser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M (2011) Global quantification of mammalian gene expression control. *Nature* **473:** 337-342

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* **13:** 2498-2504

Silva JC, Gorenstein MV, Li GZ, Vissers JP, Geromanos SJ (2006) Absolute quantification of proteins by LCMSE: a virtue of parallel MS acquisition. *Mol Cell Proteomics* **5:** 144-156

Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, Diemer K, Muruganujan A, Narechania A (2003) PANTHER: a library of protein families and subfamilies indexed by function. *Genome Res* **13:** 2129-2141

Vogel C, Chothia C (2006) Protein family expansions and biological complexity. *PLoS Comput Biol* **2:** e48

Yates JR, 3rd, Eng JK, McCormack AL (1995) Mining genomes: correlating tandem mass spectra of modified and unmodified peptides to sequences in nucleotide databases. *Anal Chem* **67:** 3202-3210